Stochastic parcel tracking in an Euler-Lagrange compartment model for fast simulation of fermentation processes

Cees Haringa[1,*], Wenjun Tang[1,2], Henk J. Noorman[1,2]

Abstract

Compartment modeling (CM) [1, 2, 3] is a well-known approach for computationally affordable, spatially-resolved hydrodynamic modeling of unit operations. Recent implementations use flow profiles based on CFD simulations [4, 5, 6], and several authors included microbial kinetics to simulate gradients in bioreactors [7, 8, 9, 10]. However, these studies relied on black-box kinetics, that do not account for intra-cellular changes and cell population dynamics in response to heterogeneous environments. In this paper, we report the implementation of a Lagrangian reaction model, where the microbial phase is tracked as a set of biomass-parcels, each linked with an intra-cellular composition vector and a structured reaction model describing their intra-cellular response to extracellular variations. A stochastic parcel tracking approach [11] is adopted, in contrast to the resolved trajectories used in prior CFD implementations [12, 13, 14]. A penicillin production process is used as a case-study [15]. We show good performance of the model compared to full CFD simulations, both regarding the extra-cellular gradients [14] and intra-cellular pool response [16], provided the mixing time in the CM matches the full CFD simulation; taking into account that the mixing time is sensitive to the number of compartments. The sensitivity of the model output towards some of the inputs is explored. The coarsest

*Corresponding author:
Email address: c.haringa@tudelft.nl (Cees Haringa)
[1]Bioprocess Engineering, Biotechnology department, Delft University of Technology, Lorentzweg 1, 2628 CJ Delft, the Netherlands
[2]DSM Biotechnology Center, Alexander Fleminglaan 1, 2613 AX Delft, the Netherlands

representative CM requires a few minutes to solve 80 hours of flow time, compared to approx. 2 weeks for a full Euler-Lagrange CFD simulation of the same case [16]. This alleviates one of the major bottlenecks for the application of such CFD simulations towards analysis and optimization of industrial fermentation processes.

Keywords: CFD, Compartment model, Euler-Lagrange, Fermentation, Metabolic modeling

---

Abbreviations

CFD	Computational Fluid Dynamics
CRD	Computational Reaction Dynamics
CM	Compartment Model
EL	Euler-Lagrange
PBM	Population Balance Model
RTD	Residence Time Distribution
SD	Scale-Down
STR	Stirred Tank Reactor
UDM	User Defined Memory
UDF	User Defined Function

Roman	Units	Description
A	m^2	Area (general)
C_s	mol/kg	Substrate concentration
C_{ext}	mol/kg	Generic extracellular concentration
C_{int}	mol/g_{dw}	Generic intra-cellular concentration
C_x	g/kg	Biomass concentration (compartment)
$C_{x,p}$	g/$parcel$	Biomass concentration (parcel)
F	kg/s	Feed rate (general)
F_s	mol/s	Substrate feed rate (general)
K_s	mol/kg	Affinity constant for substrate
N_s	mol	Substrate amount
N_c	–	Total number compartments
N_{ax}	–	Axial compartments
N_r	–	Radial compartments
N_θ	–	Tangential compartments
N_p	–	Total number parcels
N_{liq}	–	Number of liquid phase species
N_{pool}	–	Number of intra-cellular species
P_{jump}	–	Probability parcel p leaves compartment n
$P_{dest,i}$	–	Probability parcel p enters compartment i
q_p	$mol_p/Cmol_x$/h	Specific production rate of product
q_s	mol_s/Cg_{dw}/s	Specific uptake rate of substrate
$q_{s,max}$	mol_s/Cg_{dw}/s	Max. Specific uptake rate of substrate
Q_{jump}	–	jump quantifier
r_p	μmol/g_{dw}/h	Reaction rate, parcel-based
R_s	mol/s	Reaction rate of s, compartment-based
S	–	Stoichiometric matrix
t	s	Time (general)
Δt_{max}	s	Maximum timestep size
V	m^3	Tank volume
V_n	m^3	compartment volume
V_T	m^3	Total volume (general)
X_p	μmol/g_{dw}	Intra-cellular pool

Greek	Units	Description
ρ	kg/m^3	Density
μ	h^{-1}	Growth rate
Φ	m^3/s^{-1}	Flow matrix
ϕ_n	m^3/s^{-1}	Flow out of compartment n
ϕ_{ni}	m^3/s^{-1}	Flow from n to i
ψ	$-$	Uniform random number
τ_{95}	s	Mixing time
τ_{circ}	s	Circulation timescale
τ_{rxn}	s	Uptake timescale of substrate
τ_{run}	s	Simulation runtime
τ_n	s	Residence time, compartment n
ϕ_{ij}	kg/s	Flowrate from compartment i to j

1. Introduction

Assessing the impact of environmental heterogeneity in industrial fermentation processes is a challenging aspect of process development. Due to the disparity between timescales of nutrient uptake and of mixing, nutrient gradients may exist in production fermentors [1, 17] which may impact process performance [17, 18]. The poor predictability of this impact poses a scale-up risk, and may lead to under-performance of industrial fermentation processes compared to their lab-scale counterparts. As cells perceive heterogeneous environments in the form of temporal fluctuations, experimental studies rely on 'scale-down simulators' that impose temporal fluctuations from the perspective of the cell [19, 20, 21]. Ideally, such scale-down studies represent the (expected) large-scale environment, but quantifying this environment is not trivial for existing fermentors, let alone conceptual designs.

Due to limitations in experimental assessment in production fermentors, quantification of the large-scale environment generally relies on simulations, combining Computational Fluid Dynamics (CFD) with 'Computational Reaction Dynamics' (CRD). To incorporate adaptation of the microbes to fluctuating conditions, population balances [22, 23, 24] or agent-based (Lagrangian) reaction models are used [12, 13, 16], where the biomass is discretized into 'biological parcels' (parcels). But while CFD allows fermentor hydrodynamics to be studied in great detail, the computational burden constrains use of combined CFD-CRD. Even with considerable simplifications such as a frozen flowfield, days to weeks of computation time may be required to study a fed-batch process [16].

Compartment models (CM) form an interesting middle-ground between ideal reactor models and full CFD. Originally these models were based on experimental data [1, 3], nowadays CFD is typically used to compute inter-compartment flows [4, 5, 6]. Combined with black-box kinetics [7, 8, 9, 10, 25], such models provide information on large-scale gradients in seconds. However, black-box kinetics rely on the assumption of instantaneous equilibrium between intra- and extra-cellular conditions, which is questionable. As with full CFD, more realis-

tic reaction dynamics can be incorporating via population balances [24, 26] or agent-based models. While a methodology for parcel tracking in compartment models has been published previously [11], biological kinetics were not included. Hence, in this work we study the implementation of agent-based CRD in a compartment model. We focus on the technical implementation and benchmarking of the outcome against full-CFD simulations. Whilst a considerable reduction in computation time is already evident in the current implementation, further optimization in order to minimize the computation time is a subject for further study.

2. Compartment model setup

Penicillin production in a 54m^3 stirred fermentor is used as a case study; The CFD simulations underlying the compartment model have been described in prior work [14, 16]. Compartments are generated using a homogeneous cylindrical grid,[6]. In the rest of this work, the used grid is indicated as $A[N_{ax}]R[N_r]T[N_\theta]$, with N_{ax}, N_r, N_θ indicating the number of compartments in the respective dimensions. Conceptually, the methods described here are equally applicable to gradient-based compartment layouts [4, 5, 9]. Compartment generation is conducted in ANSYS FLUENT via a *User Defined Function* (UDF), which assigns a compartment number to each gridcell, sums the cell volumes within each compartment to acquire the compartment volumes, and loops over all faces on compartment interfaces to determine the inter-compartment flowrates (convective and turbulent), using the same formulation as Delafosse et al. [6]. The data is exported in text files. We opt for this UDF-based approach because it provides excellent closure of the compartment mass balances. The text files containing compartment data are imported into Julia v1.5 (https://julialang.org/), which was used for all CM-CRD calculations. The package *differentialequations.jl* [27] is used to solve the system of ODEs describing the species balances of intra- and extra-cellular components. The current model implementation considers the liquid phase and the parcel phase, resulting

in a set of $N_{tot} = N_c \cdot N_{liq} + N_p \cdot N_{pool}$ equations, with the first term representing the liquid phase balances, and the second term the intra-cellular pool balances for each parcel. In this work we focus solely on glucose as a substrate ($N_{liq} = 1$).

2.1. Reaction calculations

Using agent-based model with structured, multi-pool kinetics, the reaction kinetics are calculated for each individual parcel. The rates r_p may be a function of both the extra-cellular concentration $C_{s,n}$ in compartment n (where the parcel resides) and intra-cellular pools X_p. The differentials for intra-cellular pools are then determined by multiplication with the stoichiometric matrix S

$$dX_p = S \cdot r_p - \mu \cdot X_p \quad [\mu\text{mol/g}_x\text{/h}] \tag{1}$$

Where the term $\mu \cdot X_p$ represents pool dilution due to growth. For extra-cellular species S, the differential equations for the liquid mass balance (in moles) read:

$$dN_S = \Phi \cdot (N_c/V_c) + F_s - R_s \quad [\text{mol/s}] \tag{2}$$

With Φ the inter-compartment flux matrix (consisting of convective plus turbulent flow), V_c a compartment volume vector, F_s a feed vector (in mol/s) and R_s the uptake vector. This vector represents the coupling between cell and environment: first the consumption rate is computed for each individual parcel; uptake from a compartment n is the sum of consumption rates for the parcels residing in n:

$$R_{s,n} = \Sigma_{p \,\rightarrow\, n}(q_{s,p} \cdot C_{x,p}) \quad [\text{mol/s}] \tag{3}$$

where $q_{s,p}$ is the consumption rate in the metabolic model and $C_{x,p}$ the quantity of biomass associated with a parcel. The set of differential equations is solved with the Bogacki-Shampine 3/2 method [28] as it provides a good balance

between accuracy and speed (Appendix B). However, we note the best solver may vary depending on the problem size and stiffness.

2.2. parcel position updating

In contrast to Delafosse and Delvigne [11], who pre-compute the position vectors for each parcel, we compute positional updates within the ODE-update function. There is a benefit in terms of memory usage, as only the current position is stored, but the need to update positions every timestep introduces a small performance penalty. In future work, we may consider the comparative performance of these methods. Parcel transport consists of two steps: 1) determining whether a parcel p leaves compartment n, and 2) determining the destination compartment i in case it does (most compartments have multiple neighbors). The first is determined by the compartment residence time: $\tau_n = V_n/\Sigma\phi_n$, where $\Sigma\phi_n$ is the sum of all flowrates leaving n, convective and turbulent. The probability of parcel p leaving n in timespan Δt hence equals $P_{jump}(p,n) = 1 - \exp(-\Delta t/\tau_n)$. For step (2), the relative probability for jumping to compartment i is $P_{dest,i} = \phi_{ni}/\Sigma\phi_n$, with ϕ_{ni} the flowrate from n to i (convective *plus* turbulent). Practically, the two steps can be combined by introducing a "jump quantifier" Q_{jump} as highlighted in pseudocode-algorithm 1; if Q_{jump} is negative, the parcel stays in place. If it is positive, Q_{jump} ranges from 0 to 1 and can be used to determine the destination compartment.

Algorithm .1: Pseudocode describing the ump determination algorithm used in the CM model.

$\psi = rand(1)$ #draw uniform random number between 0 and 1
$Q_{jump} = \frac{P_{jump}(p,n) - \psi}{P_{jump}(p,n)}$ # jump quantifier
if $Q_{jump} =< 0$ not jumping
else $Q_{jump} > 0$ jump, determine destination
if $P_{dest,1} >= Q_{jump}$ jump to 1st neighbour
elseif $(P_{dest,1} + P_{dest,2}) >= Q_{jump}$ jump to 2nd neighbour
elseif $(P_{dest,1} + P_{dest,2} + ... + P_{dest,m}) >= Q_{jump}$ jump to m^{th} neighbour

We verified the implementation by confirming that the mixing time measured from particulate distribution equalled that of using an Eulerian tracer (see Appendix A).

2.3. Modeling steps

The compartment model was developed and evaluated in four steps:

- Step I: Flow implementation and mixing behavior
- Step II: Implementation of (black box) Monod kinetics
- Step III: Implementation of structured kinetics - chemostat
- Step IV: Implementation of structured kinetics - fed batch

For each step, the results are compared with prior CFD simulations [14, 16]. Besides these steps, a sensitivity study towards the accuracy in the prediction of the penicillin production rate (q_p) and the computation time is done, where the impact of several factors is studied: N_c, N_p, max. timestep size Δt_{max}, integrator relative tolerance and integration algorithm. For brevity, the sensitivity study is reported in Appendix B. We use a case with 26 axial, 6 radial and no tangential divisions ($A26R6T1$) as a base case compartment layout. Unless otherwise mentioned, $N_p = 1000$, an integrator RelTol of 0.001 is used, and $\Delta t_{max} = 0.03$.

3. Results and discussion

3.1. I: Flow and Mixing

We first confirm the capability of the CM to reproduce mixing from the Eulerian perspective. Here, the mixing time is monitored at a single probe point, equal to the CFD simulations [14]; mixing curves for selected cases are visualized in fig 1. **A**; the 95 % mixing time is visualized in fig 1. **B**. In line with Delafosse et al. [6], τ_{95} matches the CFD simulations well for a large number of compartments N_c. But, ideally we want to compute with N_c as small as possible, to minimize computational demand. In the low N_c-range, a broad scatter in τ_{95} is observed: few radial compartments N_r result in an over-estimation of τ_{95} (due to under-resolving the re-circulation), while few axial compartments

N_{ax} leads to under-estimation of τ_{95} (due to under-resolving axial mixing resistance [29, 30]). Properly balancing N_r and N_{ax} can lead to a good solution in terms of mixing time. Keep in mind that this solution cannot be considered 'grid-independent' as it is acquired by balancing out errors, but provided the prediction of the magnitude of the gradient, and subsequent metabolic calculations, are not impacted by the low N_c, this approach may provide a pragmatic, computationally manageable model of bioreactor heterogeneity. We further explore this in section 3.3. Overall we conclude the mixing features observed in the full CFD simulation are well represented in the CM model, provided N_C is chosen appropriately (or set large enough). A more elaborate Eulerian mixing verification, including comparison to lab-scale mixing data [31], is presented in Appendix A.

Next, to verify the Lagrangian implementation, the parcel mixing time is registered with $N_p = 10^3 - 10^6$ in the base case compartment layout. Mixing in the full domain is monitored with the coefficient of mixing (CoM) (eq. 4) [32]. Figure 1, **C** compares the tracer-based and parcel-based mixing curves and reveals excellent agreement for N_p set suffciently large. For lower N_p, statistical fluctuations in the local parcel concentration prohibit the threshold for 95 % mixing from being reached, but this is not necessarily problematic: provided fluctuations are fast compared to the reaction timescale (τ_{rxn}), these fluctuations do not propagate (significantly) in the reaction model. Importantly, no *systematic* offset is observed between the mixing dynamics in the Lagrangian and Eulerian perspective. In Appendix A we show some offset is observed if large timesteps are taken, as a large Δt introduces a bias towards longer residence times in the parcel-jump algorithm. However, we found that once kinetics are involved, adaptive timestepping keeps Δt sufficiently small to avoid such effects (Appendix B).

$$CoM(t) = \sqrt{\left(\frac{\Sigma_n (\frac{C_{n,t} - \overline{C}}{\overline{C}})^2 \Delta V_n}{\Sigma_n \Delta V_n} \right)} \tag{4}$$

3.2. II: Monod kinetics

Black-box Monod kinetics ($q_s = q_{s,max} \cdot (C_s/(K_s + C_s))$, with $q_{s,max} = 4.444 \cdot 10^{-7}$ mol/g_{dw}/s and $K_s = 7.8$ µmol/kg) were implemented in both an Eulerian and Lagrangian framework, to verify the predicted substrate gradient with respect to the CFD simulation reported in prior work [14]. A constant glucose feed of 1.23 g/m^3/s is added at the feed location ($y = 7.4$ m, $r = 0.75$ m). The biomass concentration is $C_x = 55$ g_{dw}/kg; in the Lagrangian implementation, this translates to $C_{x,p} = C_x \cdot (V_T \cdot \rho_l)/N_p$ gram biomass per parcel, with V_T the total reactor volume. To facilitate comparison, the excess/limitation/starvation regime definition [14] is used to visualize the gradient. Figure 2 compares the observed regime division for various compartment layouts, comparing the Eulerian and Lagrangian formulation for reaction coupling, with the full CFD simulation. As mentioned above, the finite N_p will cause some fluctuations in local biomass concentration, which in turn induce spurious substrate concentration fluctuations, and hence regime fluctuations. The Lagrangian regimes are hence averaged over 1800 s (ca. 30 mixing times). The regime distribution is reproduced to a satisfactory degree, although the size of the excess and limitation regimes is generally somewhat over-estimated in the CM (note that in all selected compartment layouts, τ_{95} was approximately but not exactly matched with CFD).

The spurious oscillations in local substrate concentrations due to the finite N_p are visualized in fig. 3 at three spatial locations; lower N_p naturally results in stronger oscillations. A consistent, positive offset in C_s can be observed, which is more pronounced at low N_p. In brief, this offset is caused by clipping of the reaction terms in case uptake exceeds availability (it is addressed in more detail in [33] in the CD context). This offset explains why, in fig. 2 and quantitatively noted in table 1, the starvation zone is somewhat smaller for all Lagrangian cases. A higher N_p will reduce the offset, but at cost of increased computation time. Whether these spurious oscillations translate into the intracellular response depends on the timescales involved; in case the timescale of

artificial extra-cellular oscillations matches the timescales related to the related intra-cellular pools, artificial intra-cellular oscillations may be observed, which could result in erroneous results (e.g. if there are irreversible or hysteresis effects in the metabolic model). In the current implementation, some intra-cellular oscillations in pool $Xgly$ are observed (especially with low N_p), however, these oscillations do not lead to consistent offsets, and the oscillations do not substantially propagate to other intra-cellular pools due to their much longer turnover times.

3.3. III: Structured kinetics: chemostat

The $9-pool$ model for penicillin production [34] was implemented to study the performance with structured kinetics. We replicate the "TU-A" CFD simulation of our prior CFD work [16]: 80 h of flow-time are simulated to establish a steady-state in intra-cellular pools, the concentration of glucose transporters is fixed in chemostat mode to achieve $q_{s,max} = 1.13 \cdot 10^{-3}$ mol/g$_{dw}$/h, and $K_s = 9.8$µmol/kg. The base-case ($A26R6T1$) simulation is compared with the CFD results in fig. 4, with $N_p = 1000$. Data was stored every 3600 s to minimize time spent on data writing, and averaged over all parcels. The results match very well between CFD and CM, which shows the successful integration of agent-based structured kinetics in the compartment model framework. In table 2, results are shown for several other compartment realizations with similar τ_{95}, which reveals q_p is well estimated even with low N_c and N_p. The advantage of using the CM over CFD is striking: while full CFD required around 1 day of computation *per hour* of flow time (with frozen flow, $\Delta t = 0.03$ and $N_p = 2500$), the CM required slightly more than an hour to run 80 *hours of flow time* with $N_p = 1000$, and less than 2 min for the coarsest realization. A wide range of cases, and sensitivity to various settings, is analyzed in Appendix B. Naturally, the numbers will be dependent on the model complexity: while low N_p facilitates a fast runtime, it is not suitable to study the potential emergence of population heterogeneity; a larger N_p is required to model the population distribution, with due computational cost. Another application of the CM-CRD

approach may be lifeline-based downscaling; figure 5 shows parcel lifelines from the above-mentioned cases. While the lower resolution removes rapid turbulent fluctuations, these lifelines clearly reflect the major fluctuations observed in full-CFD simulations [14]. As such, these lifelines can be analyzed for downscaling purposes analogous with the approaches discussed in earlier work [14, 35].

3.4. IV: Structured kinetics: fed-batch

To conclude, we compare the compartment model with the fed-batch simulation reported in prior work [16]. In the current implementation, we are subject to the same simplifications as the original CFD, particularly the assumption of a constant volume. In future work, this limitation may be lifted, e.g. by stepwise updating of the compartment volumes, as recently reported by Nadal-Rey et al. [25]. In contrast to the chemostat simulations in sec. 3.3, the amount of biomass per parcel $C_{x,p}$ and the glucose transporter concentration X_{11} are now dynamic pools. Figure 6 shows the comparison between the CFD simulation, the CM base-case in fed-batch mode, and a very low resolution case (A18R2, $N_P = 36$, $\Delta t_{max} = 0.03$ s). In addition, ideally mixed realizations with a black-box model [16] and the 9-pool model are added. Fig. 6 **A** provides the imposed feed profile, which is corrected for the constant volume assumption that has been applied in the current CM as well as the previous CFD simulation. Fig. 6 **B** shows the mean biomass concentration, computed as $C_x = \Sigma(C_{x,p})/(\rho_l V_T)$. The profiles from both CM simulations very tightly follow the CFD result; due to the stochasticity induced by the low number of parcels, the growth rate (fig. 6 **C**) is more strongly oscillating in the CM simulations. Fig. 6 **D** shows the q_p is in very good agreement with the CFD simulations, too. If we consider the intra-cellular pools (fig. 6 **E**), we see equally good agreement between the CM and CFD simulations, with only minor offsets in mean pool size and the standard deviation, the latter being representative of population heterogeneity. The emergence of population heterogeneity in the base case CM is further visualized in figure 6 **F**. This figure shows the distribution of the amino acid pool, glucose transporter pool and penicillin producing enzyme pool. As in fig. 6

E, we observe excellent qualitative agreement with a minor quantitative offset compared to the full CFD simulations. As for the previous full-CFD approach, the CM approach is capable of making predictions regarding the emergence of population heterogeneity, although we must stress (as in our previous work [16]) that the predictions regarding population heterogeneity have not been experimentally verified. At this point, they mostly serve to present hypotheses for experimental follow-up. We think such experimental follow-up, as well as the development of cell models explicitly designed to model (experimentally observed) population heterogeneity, presents a very interesting avenue for further work.

In previous work, we compared the CFD model with the ideally-mixed model with black-box kinetics. Currently, we also included a single-compartment (ideally mixed) realization of the 9-pool model, to indicate how much of the q_p-reduction results from spatial heterogeneity, and how much from the impact of low growth rates. Figure 6 **D** indicates the low growth rate substantially contributes to productivity loss, more so than in the black box model (prospectively due to the reducing X_{E11}, the glucose transporter pool, in the 9-pool model). While the q_p is still somewhat over-estimated compared to plant data, the conceptually simpler and faster-solving ideally mixed implementation may suffice for initial screening purposes. The CM-CRD model may then be utilized to produce a more accurate quantitative prediction, accounting for the productivity losses due to spatial heterogeneity, in a computation time that is vastly superior to the full CFD simulation.

4. Conclusions

In this work, we report a compartment model using stochastic parcels to represent the biotic phase, akin to Lagrangian tracking in CFD simulations. The introduction of a particulate phase allows the coupling of structured metabolic models, considering the microbial response in multiple intra-cellular pools. In addition, the fluctuations in extra-cellular conditions experienced by microbes

can be monitored in these compartment models, which, like their CFD counterparts, can be used in the design of scale-down simulators [35]. The use of compartment models for such studies provides a strong computational gain over full-CFD models; even with many simplifications and frozen flow, an Euler-Lagrange CFD study can take days to weeks for a full batch - compared to minutes for the coarsest compartment models. Nevertheless, the main features of extra-cellular gradients are maintained, and the microbial response with a structured kinetic model closely follows the results of the full CFD simulation. With this, the simulations also run considerably faster than real-time, which opens up a range of new applications in, for example, process optimization [16], that are well out of reach for full CFD simulations for the foreseeable future. In this paper, we focused on outlining the concept of a Lagrangian compartment model, and comparing the results with a full CFD simulation. As such, a number of simplifications were made, similar to those used in the CFD model. Future extensions may aim at improving the physical representation of fermentation processes, by including e.g. the gas phase, mass transfer, and volume changes in fed-batch processes. Such extensions have recently been shown for Eulerian compartment models [25], and can straightforwardly be translated to the current framework. Furthermore, we did not currently aim at a full numerical optimization. While the computation time is already favorable compared to CFD simulations, further gains in processing time may be possible by numerical optimization, or considering solvers that are better suited for the stochastic nature of parcel tracking than regular ODE-solvers.

5. acknowledgements

The authors thank Victor Koppejan, Wouter van Winden and Liang Wu of the DSM Biotechnology Center and Sjef Cornelissen of Centrient Pharmaceuticals for their valuable input. This work builds upon results from previous collaborations between TU Delft, DSM, Centrient Pharmaceuticals, East China University of Science and Technology (ECUST) and Shanghai Guojia Ltd. Our

collaborators at these institutions are gratefully acknowledged for their contributions.

References

[1] N. M. Oosterhuis, N. W. Kossen, Dissolved oxygen concentration profiles in a production-scale bioreactor, Biotechnology and bioengineering 26 (1984) 546–50.

[2] P. Vrábel, R. Van der Lans, Y. Cui, K. Luyben, Compartment Model Approach, Chemical Engineering Research and Design 77 (1999) 291–302.

[3] P. Vrábel, R. G. van der Lans, F. N. van der Schot, K. C. Luyben, B. Xu, S.-O. Enfors, CMA: integration of fluid dynamics and microbial kinetics in modelling of large-scale fermentations, Chemical Engineering Journal 84 (2001) 463–474.

[4] F. Bezzo, S. Macchietto, C. C. Pantelides, General hybrid multizonal/CFD approach for bioreactor modeling, AIChE Journal 49 (2003) 2133–2148.

[5] F. Bezzo, S. Macchietto, C. C. Pantelides, A general methodology for hybrid multizonal/CFD models: Part I. Theoretical framework, Computers and Chemical Engineering 28 (2004) 501–511.

[6] A. Delafosse, M.-L. Collignon, S. Calvo, F. Delvigne, M. Crine, P. Thonart, D. Toye, CFD-based compartment model for description of mixing in bioreactors, Chemical Engineering Science 106 (2014) 76–85.

[7] R. Spann, K. V. Gernaey, G. Sin, A compartment model for risk-based monitoring of lactic acid bacteria cultivations, Biochemical Engineering Journal 151 (2019) 107293.

[8] R. Spann, J. Glibstrup, K. Pellicer-Alborch, S. Junne, P. Neubauer, C. Roca, D. Kold, A. E. Lantz, G. Sin, K. V. Gernaey, U. Krühne, CFD predicted pH gradients in lactic acid bacteria cultivations, Biotechnology and Bioengineering 116 (2019) 769–780.

[9] T. Tajsoleiman, R. Spann, C. Bach, K. V. Gernaey, J. K. Huusom, U. Krühne, A CFD based automatic method for compartment model development, Computers and Chemical Engineering 123 (2019) 236–245.

[10] G. Nadal-Rey, D. D. McClure, J. M. Kavanagh, S. Cornelissen, D. F. Fletcher, K. V. Gernaey, Understanding gradients in industrial bioreactors, Biotechnology Advances (2020) 107660.

[11] A. Delafosse, S. Calvo, M.-L. Collignon, F. Delvigne, M. Crine, D. Toye, Euler–Lagrange approach to model heterogeneities in stirred tank bioreactors – Comparison to experimental flow characterization and particle tracking, Chemical Engineering Science 134 (2015) 457–466.

[12] A. Lapin, D. Müller, M. Reuss, Dynamic Behavior of Microbial Populations in Stirred Bioreactors Simulated with EulerLagrange Methods: Traveling along the Lifelines of Single Cells, Industrial Engineering Chemistry Research 43 (2004) 4647–4656.

[13] A. Lapin, J. Schmid, M. Reuss, Modeling the dynamics of E. coli populations in the three-dimensional turbulent field of a stirred-tank bioreactor—A structured–segregated approach, Chemical Engineering Science 61 (2006) 4783–4797.

[14] C. Haringa, W. Tang, A. A. T. Deshmukh, J. Xia, M. Reuss, J. J. Heijnen, R. F. R. Mudde, H. H. J. Noorman, Euler-Lagrange computational fluid dynamics for (bio)reactor scale down: An analysis of organism lifelines, Engineering in Life Sciences 16 (2016) 652–663.

[15] G. Wang, B. Wu, J. Zhao, C. Haringa, J. Xia, J. Chu, Y. Zhuang, S. Zhang, J. J. Heijnen, W. van Gulik, A. T. Deshmukh, H. J. Noorman, Power Input Effects on Degeneration in Prolonged Penicillin Chemostat Cultures: A Systems Analysis at Flux, Residual Glucose, Metabolite and Transcript Levels, Biotechnology and Bioengineering (2017).

[16] C. Haringa, W. Tang, G. Wang, A. T. A. Deshmukh, W. A. W. van Winden, J. Chu, W. W. M. van Gulik, J. J. J. Heijnen, R. R. F. Mudde, H. J. H. Noorman, Computational fluid dynamics simulation of an industrial P. chrysogenum fermentation with a coupled 9-pool metabolic model: Towards rational scale-down and design optimization, Chemical Engineering Science 175 (2018) 12–24.

[17] G. Larsson, M. Törnkvist, E. S. Wernersson, C. Trägårdh, C., H. Noorman, S. O. Enfors, Substrate gradients in bioreactors: origin and consequences, Bioprocess Engineering 14 (1996) 281–289.

[18] S.-O. Enfors, M. Jahic, A. Rozkov, B. Xu, M. Hecker, B. Jürgen, E. Krüger, T. Schweder, G. Hamer, D. O'Beirne, N. Noisommit-Rizzi, M. Reuss, L. Boone, C. Hewitt, C. McFarlane, A. Nienow, T. Kovacs, C. Trägårdh, L. Fuchs, J. Revstedt, P. Friberg, B. Hjertager, G. Blomsten, H. Skogman, S. Hjort, F. Hoeks, H.-Y. Lin, P. Neubauer, R. van der Lans, K. Luyben, P. Vrabel, Å. Manelius, Physiological responses to mixing in large scale bioreactors, Journal of Biotechnology 85 (2001) 175–185.

[19] A. R. Lara, E. Galindo, O. T. Ramírez, L. A. Palomares, Living With Heterogeneities in Bioreactors: Understanding the Effects of Environmental Gradients on Cells, Molecular Biotechnology 34 (2006) 355–382.

[20] P. Neubauer, S. Junne, Scale-down simulators for metabolic analysis of large-scale bioprocesses., Current opinion in biotechnology 21 (2010) 114–21.

[21] G. Wang, W. Tang, J. Xia, J. Chu, H. J. Noorman, W. M. van Gulik, Integration of microbial kinetics and fluid dynamics toward model-driven scale-up of industrial bioprocesses, Engineering in Life Sciences 15 (2015) 20–29.

[22] J. Morchain, J.-C. Gabelle, A. Cockx, Coupling of biokinetic and population balance models to account for biological heterogeneity in bioreactors, AIChE Journal 59 (2013) 369–379.

[23] J. Morchain, M. Pigou, N. Lebaz, A population balance model for bioreactors combining interdivision time distributions and micromixing concepts, Biochemical Engineering Journal (2016).

[24] M. Pigou, J. Morchain, Investigating the interactions between physical and biological heterogeneities in bioreactors using compartment, population balance and metabolic models, Chemical Engineering Science 126 (2014) 267–282.

[25] G. Nadal-Rey, D. D. McClure, J. M. Kavanagh, B. Cassells, S. Cornelissen, D. F. Fletcher, K. V. Gernaey, Development of dynamic compartment models for industrial aerobic fed-batch fermentation processes, Chemical Engineering Journal (2021) 130402.

[26] M. Pigou, J. Morchain, P. Fede, M.-I. Penet, G. Laronze, An assessment of methods of moments for the simulation of population dynamics in large-scale bioreactors, Chemical Engineering Science 171 (2017) 218–232.

[27] C. Rackauckas, Q. Nie, Differentialequations.jl–a performant and feature-rich ecosystem for solving differential equations in julia, Journal of Open Research Software 5 (2017).

[28] P. Bogacki, L. F. Shampine, A 3 (2) pair of Runge-Kutta formulas, Applied Mathematics Letters 2 (1989) 321–325.

[29] C. Haringa, R. Vandewijer, R. F. Mudde, Inter-compartment interaction in multi-impeller mixing: Part I. Experiments and multiple reference frame CFD, Chemical Engineering Research and Design 136 (2018) 870–885.

[30] C. Haringa, R. Vandewijer, R. F. Mudde, Inter-compartment interaction in multi-impeller mixing. Part II. Experiments, sliding mesh and large Eddy simulations, Chemical Engineering Research and Design 136 (2018) 886–899.

[31] M. Jahoda, M. Moštěk, A. Kukuková, V. Machoň, CFD Modelling of Liquid Homogenization in Stirred Tanks with One and Two Impellers Using Large

Eddy Simulation, Chemical Engineering Research and Design 85 (2007) 616–625.

[32] H. Hartmann, J. J. Derksen, H. E. A. van den Akker, Mixing times in a turbulent stirred tank by means of LES, AIChE Journal 52 (2006) 3696–3706.

[33] C. Haringa, H. J. H. H. J. Noorman, R. F. R. Mudde, Lagrangian modeling of hydrodynamic–kinetic interactions in (bio)chemical reactors: Practical implementation and setup guidelines, Chemical Engineering Science 157 (2017) 159–168.

[34] W. Tang, A. A. T. Deshmukh, C. Haringa, G. Wang, W. van Gulik, W. van Winden, M. Reuss, J. J. Heijnen, J. Xia, J. Chu, H. H. J. Noorman, A 9-pool metabolic structured kinetic model describing days to seconds dynamics of growth and product formation by Penicillium chrysogenum, Biotechnology and Bioengineering 114 (2017) 1733–1743.

[35] C. Haringa, A. T. Deshmukh, R. F. Mudde, H. J. Noorman, Euler-Lagrange analysis towards representative down-scaling of a 22m3 aerobic S. cerevisiae fermentation, Chemical Engineering Science 170 (2017) 653–669.

Table 1: Comparison of mean extra-cellular substrate concentration and the regime division for different N_p; other settings equal that of the base-case. The Lagrangian results are averaged over 1800 s, the margin indicates 2 standard deviations.

	$N_p = 100$	$N_p = 1000$	$N_p = 10000$	CM (Euler)	CFD (Euler)
$\overline{C_s}$	$4.33 \cdot 10^{-5}$	$4.40 \cdot 10^{-5}$	$4.43 \cdot 10^{-5}$	$4.17 \cdot 10^{-5}$	$3.44 \cdot 10^{-5}$
Excess	$9.54 \pm 4.67\%$	$9.50 \pm 2.15\%$	$9.23 \pm 0.73\%$	8.68%	6.8%
Limit.	$39.0 \pm 10.8\%$	$36.3 \pm 4.26\%$	$36.5 \pm 2.02\%$	33.9%	36.2%
Starv.	$51.5 \pm 12.2\%$	$54.3 \pm 5.00\%$	$54.3 \pm 2.17\%$	57.4%	57.0%

Table 2: Performance of various compartment realizations in predicting q_p in chemostat configuration. In all cases, the total flowtime is 80 h, $\Delta t_{max} = 0.03$ s, RelTol = 0.001. More cases are reported in Appendix B.

Compartments	parcels	run-time (s)	$q_p(end)$	$\%\ diff.\ CFD$
CFD	2500	$> 10^6$	$2.99 \cdot 10^{-4}$	–
$A26R6T1$	1000	4314	$2.93 \cdot 10^{-4}$	−2.14
$A48R10T6$	1000	16601	$2.99 \cdot 10^{-4}$	0.16
$A18R2T1$	1000	4158	$2.96 \cdot 10^{-4}$	−0.92
$A18R2T1$	36	100	$2.98 \cdot 10^{-4}$	−0.37

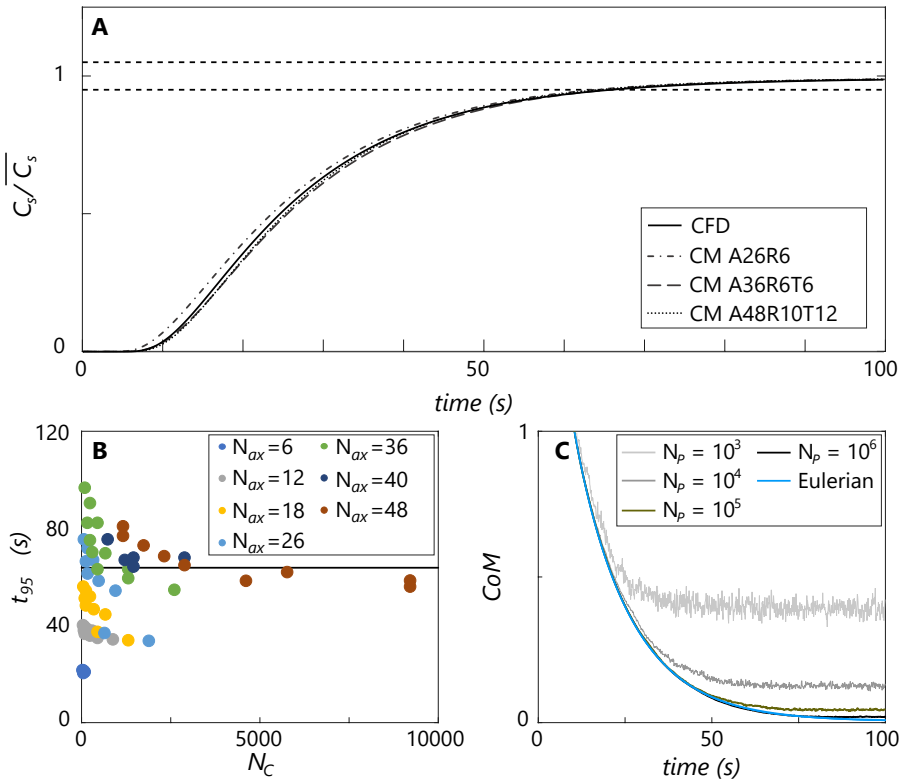

Figure 1: Mixing behavior in the compartment model. **A:** Mixing curve (single-point measurement) for several realizations of the compartment model compared to the CFD result. Note the cases were selected based on agreement with the CFD simulation; more cases are discussed in Appendix A. **B:** Point-based mixing time versus number of compartments (colored by N_{ax}) for the full range of tested compartment realizations. **C:** Comparison between Eulerian (tracer-based) and Lagrangian (parcel-based) mixing time, using the coefficient of mixing, for various N_p. $N_c = 156$ in these cases ($A26R6T1$).

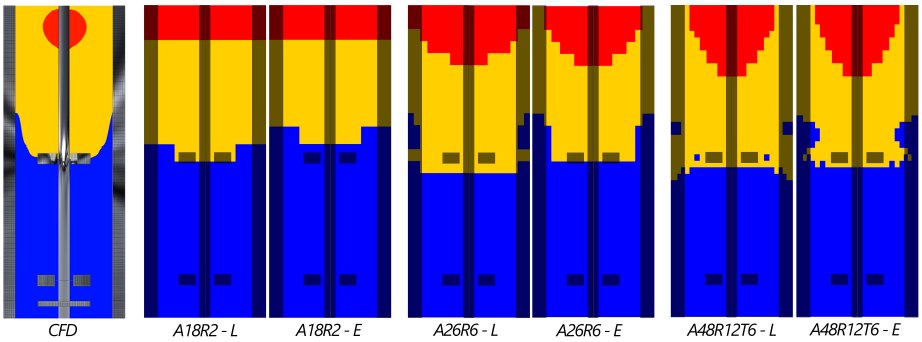

Figure 2: Regime analysis in the compartment model using Monod kinetics. Left: Full CFD model ($\tau_{95} = 63.8$). Compartment models layouts from left to right: A18R2 ($N_p = 1000$, $\tau_{95} = 56.2$), A26R6 ($N_p = 1000$, $\tau_{95} = 61.6$), A48R12T6 ($N_p = 5000$, $\tau_{95} = 68.2$)). "L" indicates the time-averaged Lagrangian result, "E" the Eulerian result.

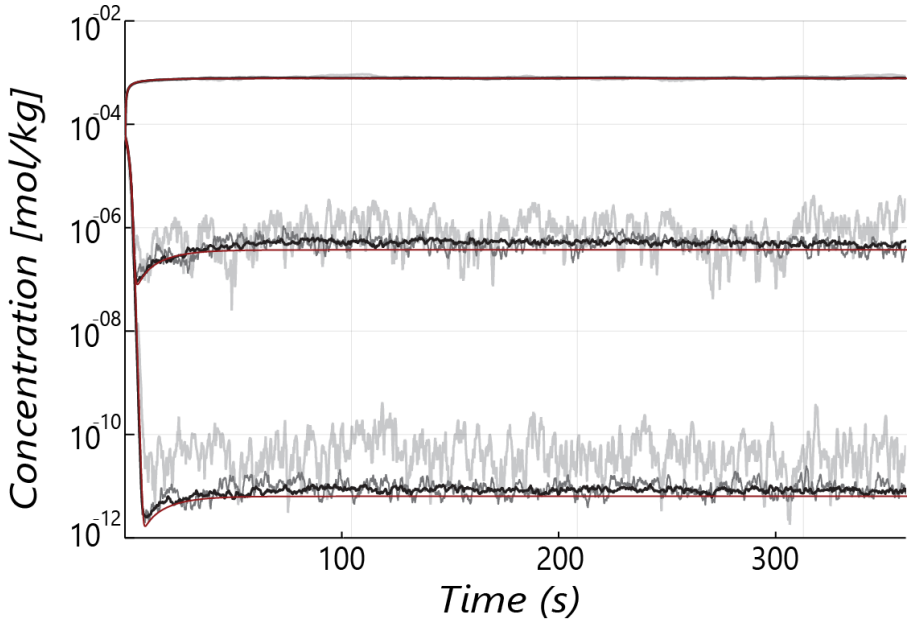

Figure 3: Oscillations in substrate concentration, measured at three different axial locations: at the feed point (top lines), at $r = 0.75$ m, $y = 3.85$ m (middle lines) and at $r = 0.75$ m, $y = 0.25$ m (bottom lines). Light gray lines: $N_p = 100$, dark gray; $N_p = 1000$, black: $N_p = 10000$. The red lines represent the Eulerian solution.

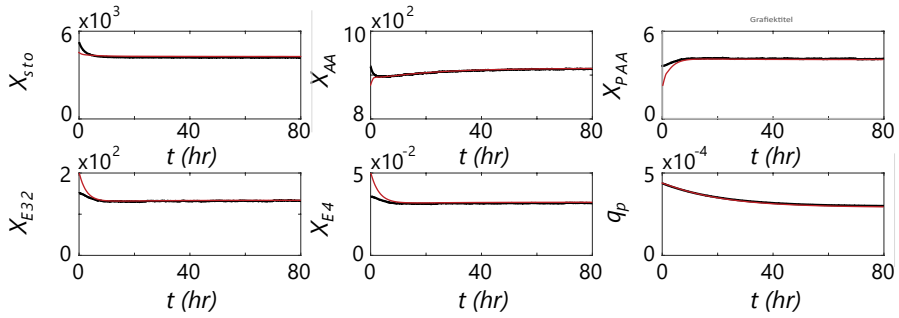

Figure 4: Pool dynamics of the CFD simulation (black) versus CM base-case (red). Offsets near $t = 0$ are caused by slightly different initial conditions; these do not affect the final 'steady' value to which the simulation converges.

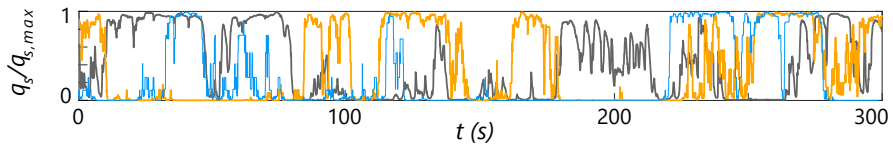

Figure 5: Example lifelines of specific glucose uptake rate (scaled with $q_{s,max}$) for a parcel in the base-case (orange) and coarse model (A18R2, blue), using the chemostat setup, compared to the CFD case (dark gray). Even with a coarse compartment model, the spatial resolution of lifelines is sufficient to reveal major fluctuations, which can be subjected to the statistical analysis described in [14, 35]

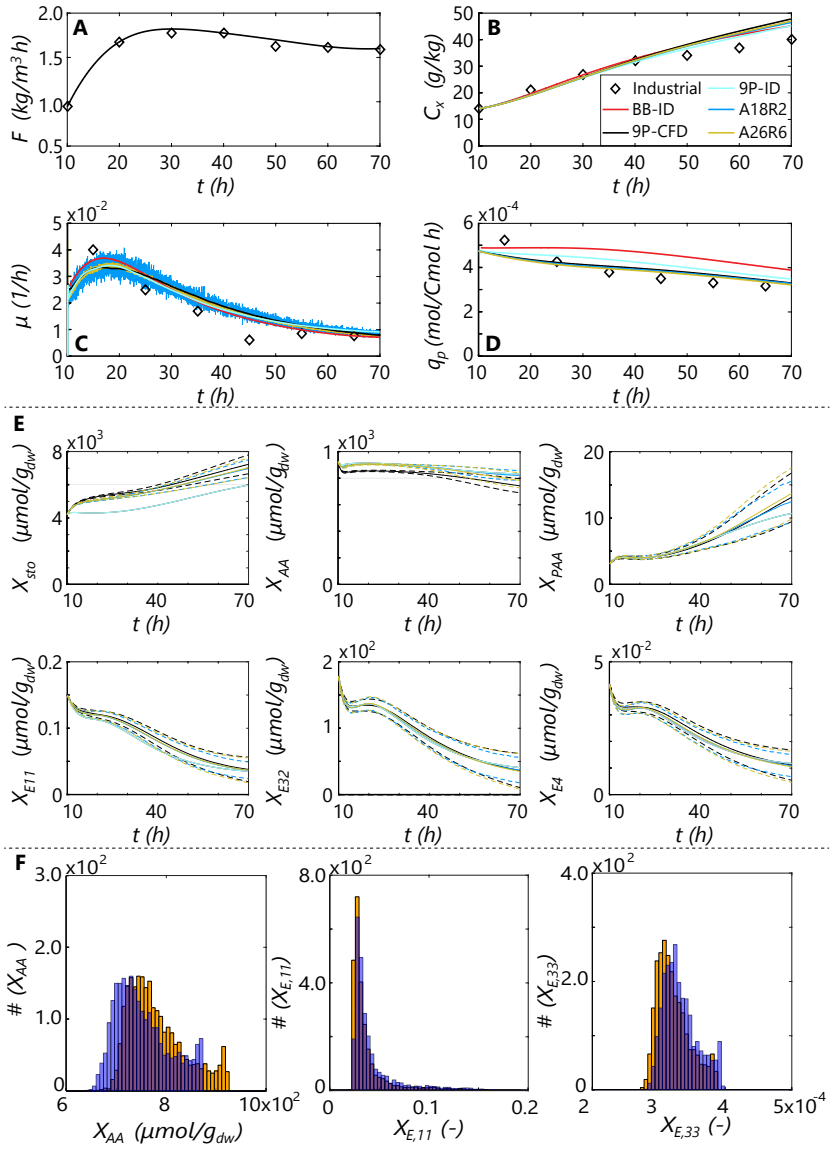

Figure 6: Dynamics of a fed-batch simulation for 2 compartment realizations (base-case and A18R2 with $N_p = 36$), compared with the full CFD simulation, an ideal-mixed model, and plant data [16]. **A:** Imposed feed profile. **B:** Biomass concentration. **C:** Growth rate . **D:** Specific penicillin production rate. **E:** parcel-averaged intra-cellular pools. The dashed lines indicate $\pm$ 1 standard deviation. **F:** Histograms of 3 pools (amino acids, transporter enzymes, penicillin production enzymes) showing population heterogeneity at the end of the process. Orange: current simulation. Transparent blue: CFD simulation from prior work [16].

Appendix A. Mixing time verification

The hydrodynamic performance of the compartment model was verified by comparing the mixing behavior observed in three stirred reactor configurations with full CFD simulations of the same systems. The following three test cases were studied:

- single Rushton impeller, full domain.
- dual Rushton impeller, 180^o domain
- dual Rushton impeller (penicillin production), 180^o domain

Note that all cases studied here have the simplification of single phase flow. We expect the methods to be equally applicable to multiphase flow. The flow patterns in the CFD simulation would be different in that case, but they can equally be translated to the appropriate inter-compartment flow maps.

Appendix A.1. Single impeller geometry

The single Rushton case is based on the geometry of Jahoda et al [31]: $T = 0.29$ m, $H = T$, $C = T/3$ and $D = T/3$. The flow is solved in steady state with the multiple-reference frame (MRF) method ($N_s = 300$ RPM), *realizable* $k - \epsilon$ turbulence model, and 2^{nd} order discretization. Since we are interested in the performance of the CM compared to CFD, no mesh dependency study was done. After flow convergence, the flow is frozen, and species tracking equations enabled. The simulation is switched to transient mode ($\Delta t = 0.01$ s). The tracer concentration development is monitored at a single probe point ($r = 9.5T/20, Y = T/4, \theta = 45^o$), the tracer pulse is injected at $t = 0$ at $r = T/4, Y = 0.28$ m, $\theta = 225^o$, both in line with the experimental setup. With this, the CFD mixing curve, reported in fig. B1, **C** is acquired.

The CFD domain is translated into a range of compartment layouts, with the following divisions: $N_{ax} = 3 - 48$, $N_r = 2 - 12$, $N_\theta = 6 - 18$. CM-mixing simulations using the same injection and detection locations are conducted in Julia. The observed τ_{95} for all compartment layouts is shown in fig. B1, **A**,

with the dots colored by N_{ax}, and the black line representing the CFD benchmark. It is visible that, while slightly underestimating τ_{95}, the compartment mixing time becomes 'grid independent' for $N_C > 2500$, within the space of chosen compartment layouts. For low N_{ax}, the mixing time is strongly underestimated because axial transport is not properly resolved, while a high N_{ax} with low N_r leads to a strong over-estimation of the mixing time, because of poor resolution of the circulation loop. N_θ has little influence on τ_{95}. As shown in fig. B1, **C**, the local dynamics of mixing are not very well captured by the CM, even if τ_{95} is comparable - the relatively large volume of compartments, modeled as ideally mixed zones, inherently induce some 'numerical blending' which affect the resolution of local dynamics. This is a conclusion similar to that by Delafosse et al. [6], even though in their 2 impeller setup, local dynamics were less pronounced to begin with, due to the resistance towards mixing being in the inter-compartment region [30, 29].

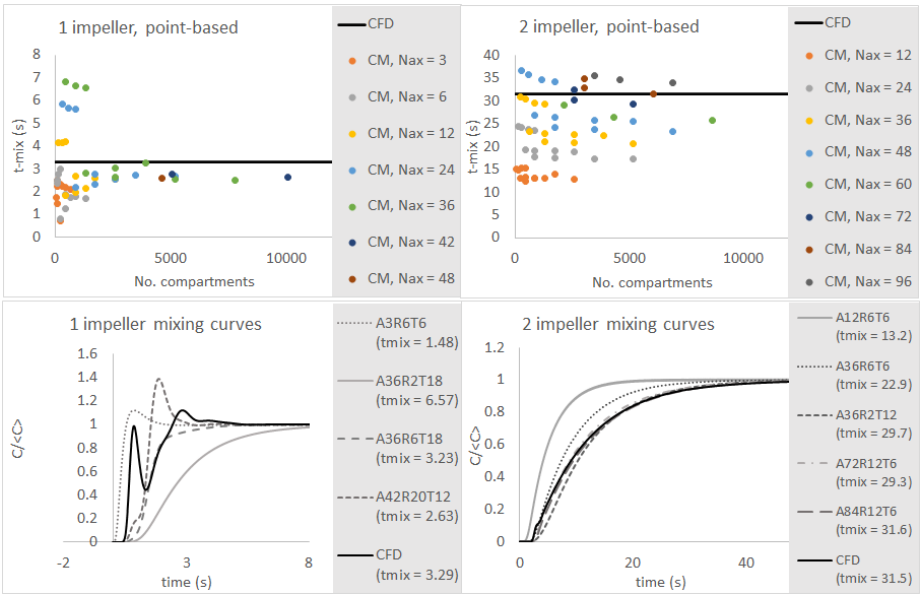

Figure B1: **A:** Mixing time (monitor point) for the 1-impeller setup of Jahoda; CFD versus various compartment realizations. **B:** As A, for the 2 impeller setup. **C:** Mixing curves for selected compartment realizations vs. CFD, 1 impeller. **D:** As C, for the 2 impeller setup.

Appendix A.2. Dual impeller geometry

The two-impeller geometry is equally based on Jahoda et al. [31]: $T = 0.29$ m, $H = 2T$, $C = T/3$, $\Delta C = T$ and $D = T/3$. The CFD setup equals that of the single impeller case, except only 180^o of the domain is modeled, with a periodic boundary condition. In this case,the monitor point was placed $r = 9.5T/20, Y = T/4, \theta = 45^o$ and a tracer pulse was injected at $t = 0$ at $r = T/4, Y = 0.57$ m, $\theta = 225^o$, yielding the mixing curve reported in fig. B1, **D**.

As for the single impeller, a wide range of compartment resolutions is employed: $N_{ax} = 12 - 96$, $N_r = 2 - 12$, $N_\theta = 3 - 18$. N_{ax} is chosen such that the impellers and inter-impeller midplane, where fully radial flow is anticipated, consistently align with compartment interfaces. The same applies to the impeller blades (3 blades in the 180^o domain) with respect to tangential divisions. Fig. B1, **B** shows the 95% mixing time, with symbols colored by the number of axial compartments. The impact of N_{ax} is even more pronounced than for the single impeller case, and as such, a wider scatter in τ_{95} vs. N_c is observed. This scatter may seem surprising, considering the good agreement observed by Delafosse et al. [6], but it must be kept in mind all their simulations featured high N_c $(4360 - 17600)$, whereas we explore much lower resolutions. With $N_{ax} > 60$, we observe results consistent with Delafosse. As before, a low N_r leads to a higher τ_{95}, although this effect is less pronounced, because the effect of low N_{ax} tends to dominate. Still, a proper 'balance' of N_{ax} and N_r can be used to yield a good prediction of τ_{95} at relatively low N_C. For example, $36 \cdot 2 \cdot 12$ compartments leads to a good prediction of τ_{95}, including local dynamics that are well in line with the CFD result (fig. B1, D), because the impact of under-resolved axial and radial flow balances out. In general, the qualitative agreement between the mixing curves from CM and CFD (fig. B1, D) at the bottom of the reactor is much better for the 2-impeller case, due to the resistance towards mixing being in the inter-impeller region, leading to more gradual mixing in the bottom region which lacks the sharp oscillations observed in the 1-impeller case. This qualitative agreement is also well in line with the observations of Delafosse

et al. [6].

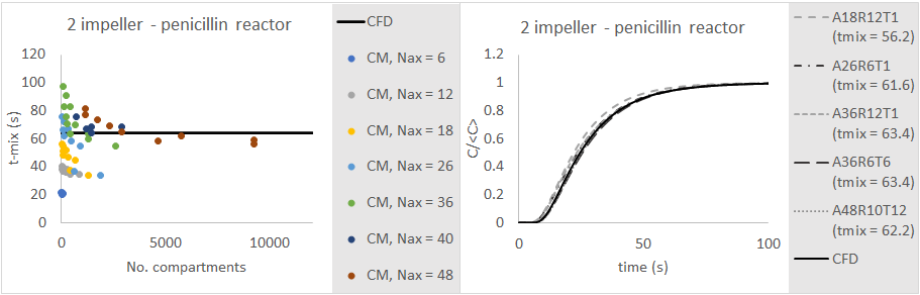

Figure B2: **A:** Mixing time (monitor point) for the penicillin reactor; CFD versus various compartment realizations. **B:** Mixing curves for selected compartment realizations vs. CFD.

Appendix A.3. Penicillin reactor

The penicillin reactor is also a 2-Rushton impeller geometry, although it differs somewhat from Jahoda et al. Most notably, the bottom impeller has 8 rather than 6 blades. The tank dimensions are $H = 7.7$ m, $C = T/3$, $\Delta C = T$ and $D = T/3$. The CFD setup used in this study is described in detail in earlier work [14]. Considering both τ_{95} versus N_C (fig. B2, **A**), and the mixing curves (fig. B2, **B**) for selected cases the trend is very similar to the Jahoda 2-impeller setup. As noted previously, this implies good results for τ_{95} can be acquired with relatively few compartments, provided the error induced by under-resolving N_{ax} and N_r balances. While this solution is not 'grid-independent', it may be a pragmatic way to reduce the required number of compartments (hence computation time) in compartment model simulations.

Appendix A.4. Mixing performance: conclusions

Overall, we observe that for a sufficient number of compartments (typically N_c approaching 10^4) the mixing behavior in the CM is close to that of a full CFD simulation, and becomes independent of N_c. At high N_c, the agreement becomes comparable to the observations of Delafosse et al. [6], although in contrast to their work, we did not explore $N_c > 10000$, considering such compartment numbers would be impractical in later metabolic simulations. For a low number

of compartments, there is substantial scatter in τ_{95} versus N_c, with under-resolution in the axial direction (low N_{ax}) leading to lower τ_{95} due to faster axial transport, and under-resolution in the radial direction (low N_r) leading to higher τ_{95} due to poor representation of the circulation flow. Properly setting the combination of N_r and N_{ax} can lead to a good prediction of τ_{95} at low N_c, which may facilitate metabolic computations by keeping the computational burden low. This result is not 'grid independent' - a poor selection of N_r and N_{ax} can easily lead to erroneous predictions in both mixing and metabolic response (more on this in Appendix B), and hence mixing performance should be validated against the CFD data. Additionally, it is recommended to verify the metabolic computations against a case with equal τ_{95} but higher N_C - even at equal τ_{95}, spatial under-resolution of the substrate gradient may affect the model output, which should be avoided.

Appendix A.5. Parcel mixing

As shown in the main text, for high N_p the Eulerian and Lagrangian mixing behavior are in full agreement. The timestep used in positional updating can have an impact on the accuracy of the Lagrangian trajectories, however. In this work, the position is updated during each function call of the integrator (which is multiple times per timestep in the multi-step methods employed here). If the Δt between two function evaluations is large, this will introduce inaccuracies in parcel behavior by introducing bias towards longer residence times. This effect is shown in figure B3, **A**. Here, simulations were conducted with various values for Δt_{max}. In addition, the Eulerian tracer concentration was set to 0 everywhere, so that all differentials dN/dt were equally zero; this results in the adaptive timestep solver (and hence parcel position updating) defaulting to the maximum stepsize. The impact of the introduced bias is clearly visible: mixing slows down compared to the Eulerian approach, and a plateau is reached at a higher CoM-value, indicating that the parcel distribution remains heterogeneous. Figure B3 **B** shows the same cases, but in this case the solution is simultaneously computing an Eulerian and Lagrangian mixing problem. In this

case, the adaptive timestepping kicks in to control the accuracy of the Eulerian approach; the timesteps become sufficiently small to accurately resolve mixing both in the Eulerian and Lagrangian approach (although statistical fluctuations in local N_p still lead to a finite plateau-CoM). These results indicate that as long as Eulerian dynamics are present, the adaptive timestep algorithm is sufficient to provide accurate Lagrangian mixing, too. Conceptually, this is logical, considering that the physical mixing timescales for the Eulerian and Lagrangian phase are similar. However, in cases where Eulerian and Lagrangian behavior are decoupled, additional care may be needed to properly set the particle position update interval. As a first estimate, $\Delta << \tau_{res}$ seems a reasonable choice, with τ_{res} the residence time in compartment with the shortest residence time.

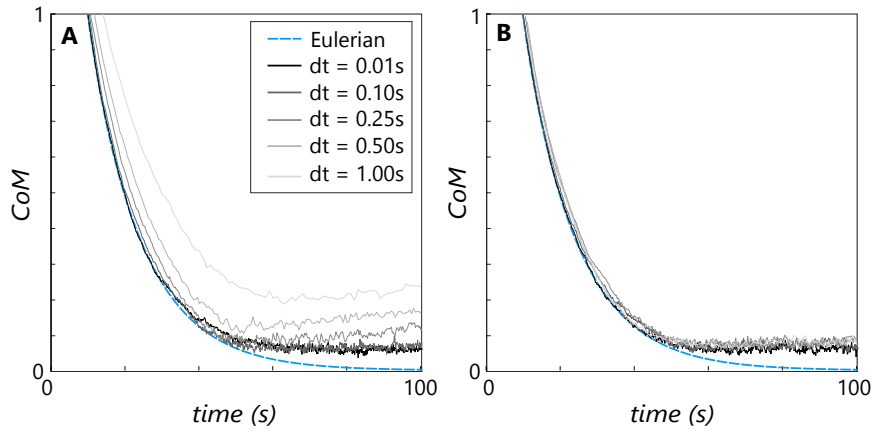

Figure B3: Lagrangian mixing behavior with $N_p = 25000$, $A18R2$, for various Δt. **A:** No timestep control. **B:** Timestep control by the Eulerian phase.

$$CoM(t) = \sqrt{\left(\frac{\Sigma_n (\frac{C_{n,t} - \overline{C}}{\overline{C}})^2 \Delta V_n}{\Sigma_n \Delta V_n}\right)} \tag{A.1}$$

Appendix B. Impact of various CM parameters on metabolic predictions

As a base case, we take $N_{ax} = 26$, $N_r = 6$, $N_\theta = 1$, $N_p = 1000$, $\Delta t = 0.03$ s and integration relative tolerance $RelTol = 1E-3$. These settings were chosen for the manageable amount of compartments ($N_C = 156$), parcels per compartment (ca. 6.5) and solution time (ca. $1-1.5$ h for 80 h flow time).

All simulations reported in this work were conducted on a desktop computer with an INTEL i5-4460 (3.2GHz, quadcore) with 16GB 1600MHz DDR3 RAM, colloquially known as "a decent midrange gaming PC built on a PhD budget in 2015". The ratings presented below are the time required to solve the ODE-section of the code: pre-processing (memory allocation, parcel initialization and preparing the property vectors and transition matrices) and post-processing (data storage and/or plotting) are excluded. Pre-processing was found to be a minor component of the total time requirement, whereas the time for plotting depends very strongly on the number of datapoints plotted, rather than the size of the system itself. All numbers were treated as Float64 (double-precision)

The current Julia code was constructed following the efficient computation guidelines offered in the DifferentialEquations.jl documentation, keeping into account the following:

- In-place computations were used to avoid allocation of memory during computation (updating the system of equations and parcel positions required 0 allocations). All arrays were pre-allocated, and no global variables were used.
- The equations were devectorized into for-loops to yield minimal overhead in the compiled code.
- Recurring mathematical operations on constants were pre-computed (e.g. recurring divisions or exponents).
- Considering the inherent fluctuations in the solution, a high numerical accuracy was not deemed necessary: the BS3 solver was used, for offering

the highest speed. A comparison with the TSIT5 solver (default) was made.

Appendix B.1. Impact of N_p

Using the base case for all other settings, N_p is varied between $50 - 2500$. With $N_{pool} = 8$ dynamic pools, the total number of equations in the system is $N_c + N_p \cdot N_{pool}$. If we plot τ_{run} versus the total number of equations, the time requirement scales linear (fig. C 1) - hence an increase in N_p gives a proportional increase in τ_{run}. The results are quantitatively summarized in table C 1 for A26R6T1. This table also list the impact on the parameter of interest in this study, q_p, at the end of the 80 h simulation timeframe. It can be observed that in all cases, the q_p in the compartment model is slightly lower than in the CFD simulation - within 2.5 % of the CFD simulation. For different compartment layouts (A6R6T1 and A36R6T1) a similar linear impact is observed. Table C 2 lists the time requirement for the other cases.

Table C 1: Impact of N_p on runtime τ_{run}. Other settings: $\Delta t = 0.03$ s, RelTol = 0.001.

Compartments	parcels	τ_{run} $[s]$	$q_p(end)$	% $diff.$ CFD
$A26R6T1$	50	344	$2.94 \cdot 10^{-4}$	-1.64
$A26R6T1$	100	485	$2.93 \cdot 10^{-4}$	-1.96
$A26R6T1$	250	1063	$2.93 \cdot 10^{-4}$	-2.08
$A26R6T1$	1000	4314	$2.93 \cdot 10^{-4}$	-2.14
$A26R6T1$	2500	10403	$2.93 \cdot 10^{-4}$	-2.09
$A26R6T1$	5000	21865	$2.93 \cdot 10^{-4}$	-2.12

Appendix B.2. Impact of N_c

In fig. C 1, **A**, the gray symbols indicate the computation time for a fixed $N_p = 1000$ and a varying number of compartments (details in table C 3). Here, a strong impact of N_C on the runtime is observed, scaling with $\tau_{run} \approx N_c^{3.9}$ in the studied range. There are a number of factors leading to this sharp increase: N_c

Table C 2: Impact of N_p on runtime τ_{run}, with different compartment numbers. Other settings: $\Delta t = 0.03$ s, RelTol = 0.001.

Compartments	N_c	N_p	τ_{run} $[s]$
$A6R6T1$	36	250	1022
$A6R6T1$	36	1000	4169
$A6R6T1$	36	2500	10262
$A36R12T1$	432	250	2110
$A36R12T1$	432	100	8293
$A36R12T1$	432	2500	19887

compartments have N_c^2 potential connections, increasing the computation time of the flow equations. In addition, the lower residence time per compartment as well as lower number of parcels per compartment (leading to larger fluctuations in local biomass concentration) increase the stiffness of the problem - requiring smaller timesteps to properly resolve parcel behavior.

As observed in table C 3, q_p differs substantially from that computed in the CFD simulation for cases A6R6T1 and A36R2T1. This offset is easy to explain: the mixing time in A6R6T1 is strongly under-estimated compared to CFD, while for A36R2T1 it is substantially over-estimated. This hence leads to a substantial under-estimation in extracellular heterogeneity for the former, and over-estimation for the latter, which affect q_p. For all the other cases, the mixing time is close to the CFD result ($\tau_{95} = 63.8$), and consequently, q_p is predicted within 2.5 % accuracy compared to the CFD result. This indicates the extracellular gradient does not need to be resolved in high spatial resolution, provided the range of C_s and the frequency at which parcels observe changes is well approximated, for which matching of τ_{mix} suffices even in coarse simulations in the current study. Of course, the impact of spatial resolution can more substantial for other metabolic models.

Table C 3: Impact of N_c on runtime τ_{run}. Other settings $N_p = 1000$, $\Delta t = 0.03$ s, RelTol = 0.001.

Compartments	N_c	τ_{run} $[s]$	τ_{95} $[s]$	$q_p(end)$	% $diff.$ CFD
$A6R6T1$	36	4169	20.3	$3.78 \cdot 10^{-4}$	26.54
$A18R2T1$	36	4158	56.2	$2.96 \cdot 10^{-4}$	-0.92
$A36R2T1$	72	4203	97.2	$2.54 \cdot 10^{-4}$	-14.92
$A26R6T1$	156	4314	61.6	$2.93 \cdot 10^{-4}$	-2.14
$A36R12T1$	432	8293	63.4	$3.00 \cdot 10^{-4}$	0.43
$A36R6T6$	1296	7549	63.3	$2.95 \cdot 10^{-4}$	-1.36
$A48R10T6$	2880	16601	65.1	$2.99 \cdot 10^{-4}$	0.17
$A48R10T12$	5760	31392	62.2	$3.03 \cdot 10^{-4}$	1.47

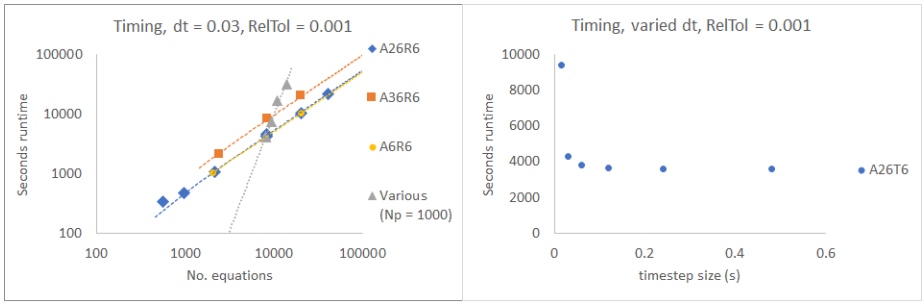

Figure C 1: Impact of various parameters on computation time. **Left:** the computation time as a function of the number of equations (parcel*pools + compartment), for 3 compartment layouts and a variable number of parcels Np. The scaling with $N_p \cdot N_{pool}$ is linear. The gray curve shows a fixed number of parcels ($N_p = 1000$), with a variable number of compartments. Scaling follows a power law N_c^n with $n > 3$, showing an increase in compartments has a large impact on computational burden. **Right:** The impact of maximum timestep size Δt on the result. For the given settings, the results indicate the maximum timestep size is not used for $\Delta t > 0.06$, hence the computation time is independent of Δt.

Appendix B.3. Impact of Δt_{max}

Fig. C 1, **B** shows the impact of Δt_{max} for the base case setup. Note that we are using a variable timestep solver, so in order to control accuracy, the used timestep may be smaller than Δt_{max}. This is clearly reflected in the figure: for $\Delta t_{max} > 0.06$ s the runtime τ_{run} is constant, indicating the maximum timestep size is not used. Conversely, the computation time for $\Delta t_{max} = 0.015$ s is double that of $\Delta t_{max} = 0.03$ s, which indicated the maximum timestep size *is* limiting the computation here (and the runtime scales inversely with Δt_{max} in this range). As mentioned in section Appendix B.2, the 'optimal' Δt likely depends on N_c because the compartment residence time decreases with an increase in compartment number, and smaller steps are needed to properly resolve parcel behavior. The results are quantitatively summarized in table C 4. Considering the agreement for q_p, a higher Δt_{max} seems to lead to a very slightly lower q_p, because there is some bias towards longer compartment residence times (hence a slight increase in parcel circulation time). However, the offset in q_p never exceeds 2.5 %, indicating the use of an adaptive timestepping algorithm prevents a residence time bias once metabolic coupling is introduced, meaning manual tuning of Δt_{max} is not required.

Table C 4: Impact of Δt on runtime τ_{run}

Compartments	Δt_{max} $[s]$	τ_{run} $[s]$	$q_p(end)$	$\%\ diff.\ CFD$
$A26R6T1$	0.015	9410	$2.94 \cdot 10^{-4}$	-1.62
$A26R6T1$	0.03	4314	$2.93 \cdot 10^{-4}$	-2.14
$A26R6T1$	0.06	3818	$2.92 \cdot 10^{-4}$	-2.38
$A26R6T1$	0.12	3633	$2.92 \cdot 10^{-4}$	-2.40
$A26R6T1$	0.24	3581	$2.92 \cdot 10^{-4}$	-2.43
$A26R6T1$	0.48	3602	$2.92 \cdot 10^{-4}$	-2.43

Appendix B.4. Impact of RelTol

Interestingly, RelTol has no large impact the solution accuracy, or the computation time (table C 5) , which may appear to contradict the statement in the previous segment. Most likely, this is because the pool dynamics are slow in comparison to $\Delta t_{max} = 0.03$ s (even with the rapid extra-cellular uptake of the current microbe, $\tau_{rxn} \approx 12\Delta t_{max}$. As long as a parcel remains in place, tolerances are easily met. The situation changes if parcels are moving around, which causes discontinuities in the uptake terms due to sudden changes in the observed substrate concentration, $C_S(p)$. Hence, we hypothesize that the previously noted impact on Δt results from the frequency of parcel jumps and consequent rapid changes in uptake (as, in this case, the Lagrangian and Eulerian frameworks are coupled), and not from the solver tolerance.

Table C 5: Impact of tolerance on runtime τ_{run}

Compartments	$RelTol$	τ_{run} $[s]$	$q_p(end)$	$\%\ diff.\ CFD$
$A26R6T1$	$1E-2$	4287	$2.92 \cdot 10^{-4}$	-2.11
$A26R6T1$	$1E-3$	4314	$2.93 \cdot 10^{-4}$	-2.14
$A26R6T1$	$1E-4$	4264	$2.93 \cdot 10^{-4}$	-2.13

Appendix B.5. Impact of Solver

All simulations were conducted with the BS3 algorithm, based on Julia benchmarks for ODE solutions with high RelTol. We compared the result to the default algorithm for differential equations - Tsit5. The difference in runtime between these algorithms is approximately a factor 2 (table C 6), without a notable difference in q_p.

Table C 6: Impact of tolerance on runtime τ_{run}

Compartments	solver	τ_{run} $[s]$	$q_p(end)$	$\%\ diff.\ CFD$
$A26R6T1$	Tsit5	8400	$2.93 \cdot 10^{-4}$	-2.19
$A26R6T1$	BS3	4314	$2.93 \cdot 10^{-4}$	-2.14

Appendix B.6. Timing: conclusions

Overall, the factors that have the biggest impact on the solution timing are N_c and N_p. In particular N_C has a substantial impact on the overall computation time, probably because increasing N_C both leads to a larger system of fluid equations, and because finer timesteps are used by the adaptive timestepper in order to properly resolve parcel behavior (due to the shorter per-compartment residence time). For the studied range of compartment numbers, the predicted q_p was negligibly influenced by N_C compared to the full-CFD result, provided the mixing time matched that of the CFD simulation. The number of parcels N_p has a linear impact on the computation time in the studied range. Interestingly, the agreement in q_p was hardly impacted by N_p, even for just 50 parcels in 156 compartments - meaning only 1 in 3 compartments even contain a single parcel at any time, and C_s in each compartment is quite variable. This is likely because the timescales of the intra-cellular pools, in particular q_p, are very long compared to the timescale of the (artificial) fluctuations induced by the under-resolved biomass phase. Hence, the impact of these spurious fluctuations is buffered. Note this conclusion may not generalize to other systems, the impact of N_p should be studied on a case-to-case basis, and with respect to the parameter of interest. In addition, note that currently only the mean pool size was compared between the CFD simulation and CM simulation, and it was not checked whether the number of parcels was adequate to fully represent the population distribution of the pool (i.e. if the higher order moments were also matched). It may be that a larger N_p is needed to ensure the full pool distribution is independent of N_p, and as such, to draw conclusions on the emergence and distribution of population heterogeneity.

In this particular case q_p is the parameter of interest, and it is well predicted (within 2.5 %) even for a low N_c and N_p, provided τ_{95} is close to the CFD simulation. The results show that for low N_c and N_p a simulation of 80 h flow time can be conducted in several minutes - compared to weeks for a full CFD simulation with frozen flow. Case in point: a final run with A18R2T1, $N_p = 36$

(1 parcel/compartment), $\Delta t_{max} = 0.5$ s and all other settings at default ran in 100 s - with a q_p offset of 0.4 % compared to full CFD.